AFOSR-TR-97

0531

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	9/9/97	Final Report-Jan. 15, '94-Jan 14, '97

4. TITLE AND SUBTITLE
Exploring Rugged Energy Landscape in Large Systems

5. FUNDING NUMBERS
F496209410106
2303/FS
61102F
F49620-94-1-0106

6. AUTHOR(S)
D. Thirumalai

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
AFOSR/NL
110 Duncan Avenue, Suite B115
Washington, DC 20332-0001
Dr Berman

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Univ of Maryland
Dept of Physical Science & Technology
College Park, MD 20742-2431

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release;
distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
During this grant period we tackled three problems all related to systems with rugged energy landscape: (1) Developing replica molecular dynamics method to detect bottlenecks in energy surface; (2) Characterizing the self-organization of biomolecules in terms of sequences; (3) Taming the energy landscape for design of biomaterials.

19971021 170

DTIC QUALITY INSPECTED 4

14. SUBJECT TERMS
Energy landscape, ruggedness

15. NUMBER OF PAGES
4

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
unclassified	unclassified	unclassified	

Proposal Title: Exploring Rugged Energy Landscape in Large Systems" - Final Report

Sponsor: Air Force Office of Scientific Research

Grant Number:F496209410106

Principal Investigator: D. Thirumalai, University of Maryland, College Park, MD 20742

The general aim of the research conducted during the funding period was to develop methods to characterize the energy landscapes (locate minima and saddle points) in large systems at finite temperatures. There are numerous problems in chemistry, biology, and material science in which one is interested in the dynamics, material and properties of systems in high dimensions. Examples where this the norm include kinetics of folding of biomolecules, dynamics of the liquid to glass transition, behavior of polymers at interfaces. For all these problems the theoretical description of the various properties require the determination of the transition states in high dimensional free energy surface. During the present funding period we accomplished three important goals related to our understanding and use of systems whose underlying energy landscape is complex. Throughout our research we used examples from the kinetics of protein folding. The methods

we developed are quite general and should be applicable in the design of novel polymeric substances. The three goals that we accomplished are: (a) Development of replica molecular dynamics (RMD) to probe the minima and barrier heights in systems with a distribution of minima and maxima. All of the examples that we have considered belong to this category. (b) The use of energy landscape ideas to classify folding kinetics of proteins. (c) Design of peptide sequences that fold into a designed structure. A brief description of each follows.

Replica Molecular Dynamics: Consider a rough energy landscape. Here there are many minima that are separated from each other by barriers that are located only on long time scales. In typical simulations of large systems it is assumed that on the computational time scale adequate sampling of the conformations take place. This is the ergodic theorem. If the simulation temperature is not high enough for the barriers to be overcome then the system will become trapped in one of the minima and reliable results cannot be obtained. The first question that needs to be answered is how the existence of the multivalley structure be detected? We developed a novel simulation method, referred to as the replica molecular dynamics (RMD), in which two replicas of the system with two distinct initial conditions are generated. For each initial condition we calculate

the time average energy of individual particles assuming that the potential energy function of the system is given. If the observation time sale is long the long time value is the average energy of the given particle which is a well defined thermodynamic quantity. We introduced the energy metric which calculated the time dependence of the distance between the time averaged values of the energy of the particles starting with the two distinct replicas. We showed that for system in which the conformational space is sampled efficiently the inverse of the energy metric goes as Dt, where D is the rate of exploration of conformational space. If there is a bottleneck between the two basins of attraction then the energy metric aturates. The temperture dependence of the energy metric gives the average barrier height separating the basins of attraction. We have applied these methods to understand the dynamics of proteins. In particular we showed rigourously that there are minima and barriers rangibg from the small energy scales to the longest scales. Thus in order to understand macromolecular self-organization one has to account for motions from few angstroms length scales to the size of the entire system.

Folding Kinetics of Proteins: A concrete test of these ides is to understand the folding kinetics of proteins. Proteins are known to be governed by a complex

energy landscape. In order to understand the kinetics in terms of the underlying rough energy surface and connect to experiments it is necessary to connect the properties of the landscape to intrinsic observables of proteins. We discovered that the folding characteristics can be succintly accounted for by the two characteristic temperatures One of them is T_θ which is the temperature at which there is a conformational collapse from an extended conformation to a compact state. The other is T_F at which the protein adopts the unique native state. We showed that minimization of the difference in these two temperatures leads to a smoother landscape thus guiding the polypeptide chain into the unique native state. Conversely if the difference is large then the landscape is rough which leads to sluggish folding. The major conclusion is that all the complexities of the rough free energy surface can be folded into this one single parameter and various scenarios for folding kinetics emerge by merely tuning this difference in temperatures.

Design of Folding Sequences: If it is correct that minimization of the two characteristic temperatures leads to fast folding in a smoother free energy surface then it must be possible to design optimized sequences that would fold into a given structure i.e., one should be able to approximately solve the inverse folding problem. We set out to do this in the context of design of four helix

bundles. It has been shown experimentally that there are few candidate sequences of aminoacids that would fold into four helix bundles. We started with a simple model containing just three aminoacid residues (following experiments). Then we performed monte carlo in sequence space and monitored those sequences that yielded low values of the difference between T_θ and T_F. The result was that there were a few sequences that can yield optimized structures. We then performed a kinetics simulations to show that these sequences indeed fold rapidly. Thus by taming the rugged landscape we have come up with a design criterion for finding sequences that fold into a given structure. We believe that this method is general and applicable for designing polymeric materials on mesoscopic length scales.θ

Publications during this grant period:

1. "Dynamics in Rugged Energy Landscapes with Applications to the S-peptide and Ribonuclease A", J. E. Straub, A. B. Rashkin, and D. Thirumalai, *J. Am. Chem. Soc.* 116, 2049-2063 (1994).

2. "Theoretical Perspectives on *in vivo* Protein Folding", D. Thirumalai, in *Statistical Mechanics, Protein Structure, and Protein Substrate Interactions*', edited by S. Doniach p. 115-134 (Plenum Press, NY, 1994).

3. "Theoretical Predictions of Folding Pathways Using the Proximity Rule with Applications to BPTI", C. J. Camacho and D. Thirumalai, *Proc. Nat'l Sci.* 92, 1277 (1995).

4. "Nucleation Mechanism for Protein Folding and Theoretical Predictions for Hydrogen-Exchange Labelling Experiments", D. Thirumalai and Z. Guo, *Biopolymers* (Research Communications 37, 137-140 (1995)).

5. "Modelling Disulfide Bonds in Globular Proteins: Entropic Barriers and Pathways", C. J. Camacho and D. Thirumalai, *Proteins: Structure, Function and Genetics*, 22, 28-40 (1995).

6. "Kinetics of Protein Folding: Nucleation Mechanism, Time Scales, and Pathways", Z. Guo and D. Thirumalai, *Biopolymers*, 36, 83-102 (1995).

7. "From Minimal Models to Proteins: Time Scales for Protein Folding Kinetics", D. Thirumalai, *J. Physique* (Paris), 5, 1457-1467 (1995).

8. "Energy Landscape and Folding Mechanisms in Proteins", Z. Guo and D. Thirumalai, *In Protein Folds: a distance based approach* edited by H. Bohr and S. Brunak (CRC Press Boca Raton, FL, 1995) 233-239.

9. "A Criterion that Determines the Foldability of Protein", D. Klimov and D. Thirumalai, *Phys. Rev. Lett.* 76, 4070-4073 (1996).

10. "Kinetics of Folding of Proteins and RNA", D. Thirumalai and S. A. Woodson, *Acc. Chem. Res.* 29, 433-439 (1996).

11. "Factors Governing the Foldability of Proteins", D. K. Klimov and D. Thirumalai, *Proteins: Structure, Function, and Genetics*, 26, 411-441 (1996).

12. "Kinetics and Thermodynamics of Folding of a *de novo* Designed four Helix Bundle", Z. Guo and D. Thirumalai, *J. Mol. Biol.* 263, 323-343 (1996).